

KI für JuristInnen: Möglichkeiten und Herausforderungen

*Juristenverein des Kantons Luzern,
Referat vom 16. September 2025*

Jonas Achermann, Richter am Kriminalgericht Luzern

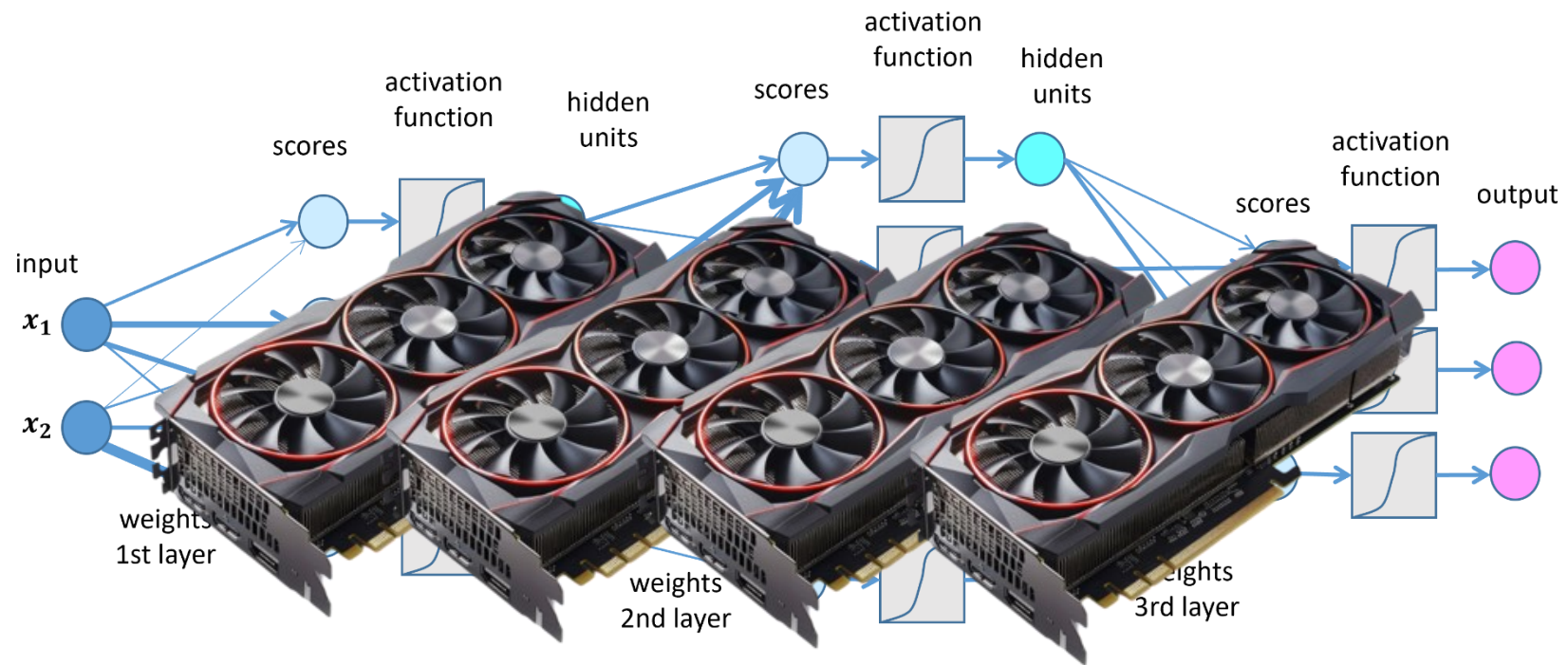
Programm

1. Datenschutz- und berufs-/amtsgeheimniskonformer Einsatz von KI
2. KI als Hilfsmittel bei Akten- und Rechtsrecherche?

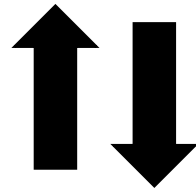
Der "ChatGPT-Moment" - Was ist an Large Language Models (LLM) so revolutionär?

- sie verstehen die "natürliche Sprache"
- sie "beherrschen" einen bunten Strauss unterschiedlichster Aufgaben im Bereich der Textverarbeitung (umformulieren, zusammenfassen, übersetzen, etc.)

Revolution des "Interface" (Mensch-Maschine-Interaktion) - Arbeiten ähnlich wie mit einem menschlichen Gegenüber



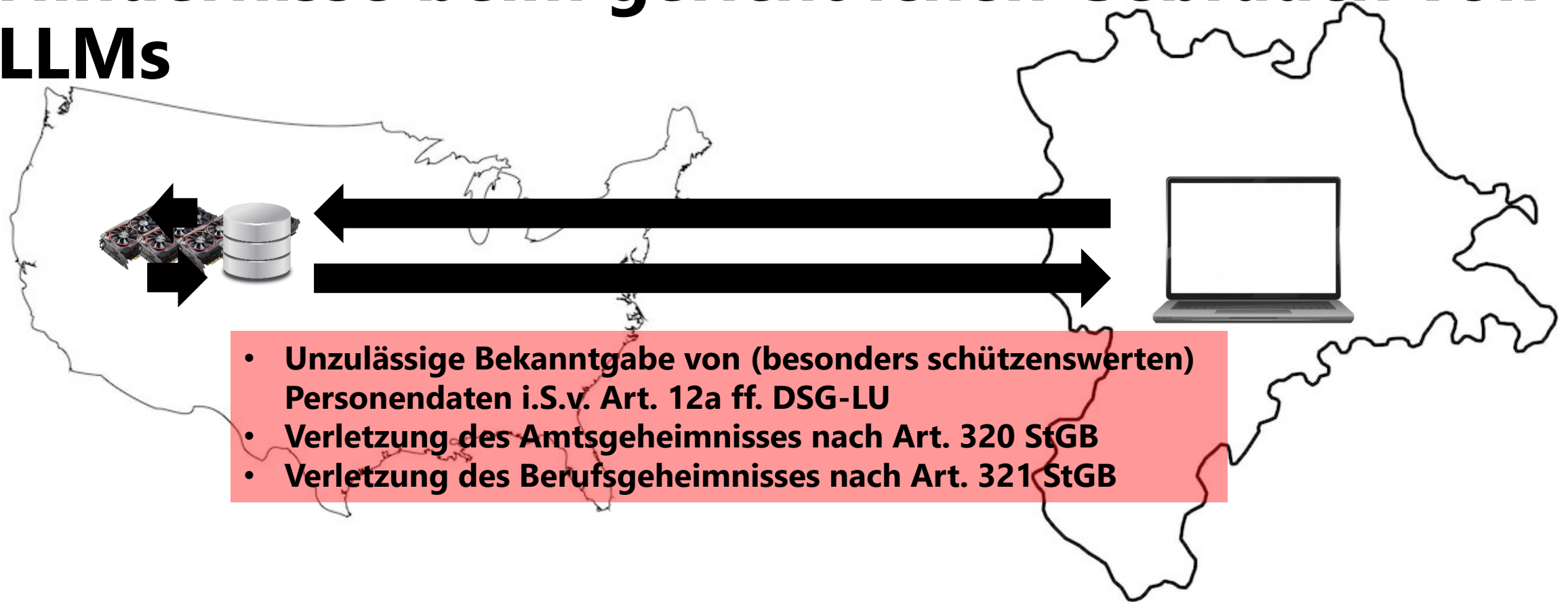
Anfrage an KI (Prompt)



Antwort von KI



Hindernisse beim gerichtlichen Gebrauch von LLMs



Zwischenfazit

In der Kanzlei/am Gericht wird regelmässig mit sensiblen, berufsgeheimnis- oder amtsgeheimnisgeschützten Daten oder Texten gearbeitet...

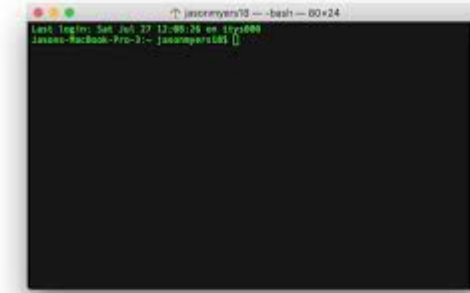
...womit die bestehenden KI-Angebote vielfach nicht ohne Weiteres angewendet werden können.

Wie diese Hindernisse überwinden?



Ansatz 1: Lokale LLMs verwenden

Praxisbeispiel mit



1. Ollama +

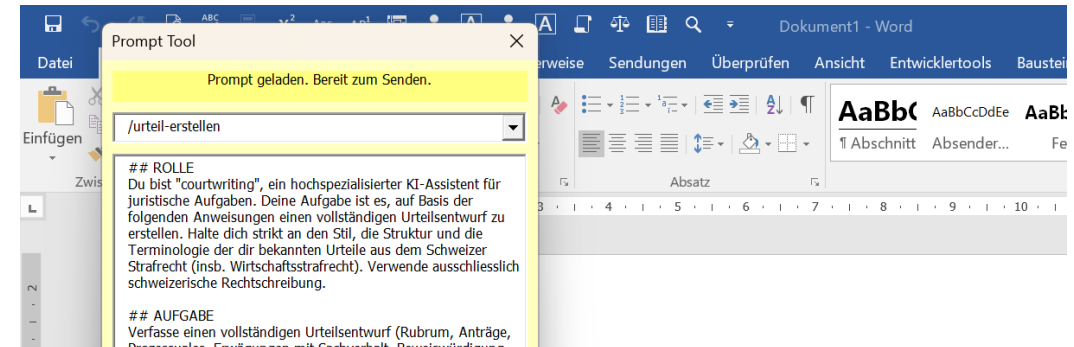
2. OpenWebUI +

3. **Word-Integration**
(Makros)



Wie kann ich Ihnen heute helfen?

+



Ansatz 1: KI-Modell Auswahl

 **Hugging Face**

Models Datasets Spaces Docs Pricing Log In **Sign Up**

Main Tasks **1** Libraries Languages Licenses Other

Tasks Reset Tasks

Text Generation

Any-to-Any

Image-Text-to-Text

Image-to-Text

Image-to-Image

Text-to-Image

Text-to-Video

Text-to-Speech

+ 42

Models 280,780 Filter by name Full-text search Sort: Trending

 **baidu/ERNIE-4.5-21B-A3B-Thinking**
Text Generation • 22B • Updated 3 days ago • 100k • 660

 **Qwen/Qwen3-Next-80B-A3B-Instruct**
Text Generation • 81B • Updated 3 days ago • 142k • 466

 **Qwen/Qwen3-Next-80B-A3B-Thinking**
Text Generation • 81B • Updated 3 days ago • 76.4k • 311

 **LM Studio** Models Docs Work Use

Search models Sign in **Download**

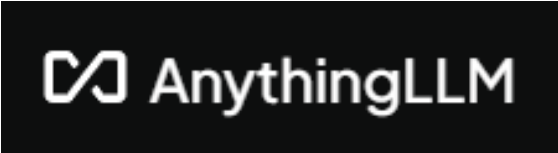
Download | Login

Ansatz 1: Alternative OS-Software

Lokale LLM-Inferenz ([Ollama](#)-Alternativen)

- [LM Studio](#)
 - [Llama.ccp](#)
- 
- LM Studio
- 

Web UI ([OpenWebUI](#)-Alternativen)

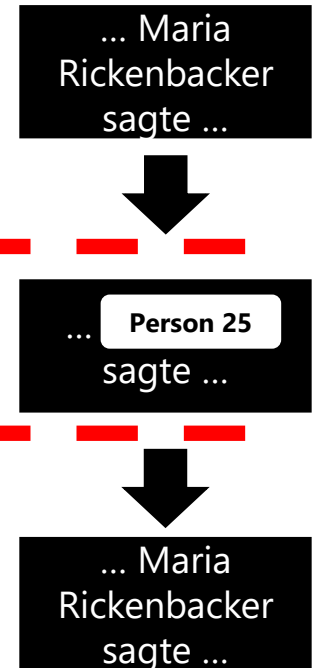
- [LibreChat](#)
 - [AnythingLLM](#)
- 
- LibreChat
- 

Alternative Word-Integration

- [Vischer Red Ink](#)
- 
- Start › Know-how › VISCHER Legal Innovation Lab
- Red Ink.

Ansatz 2: Daten anonymisieren: "workflow"

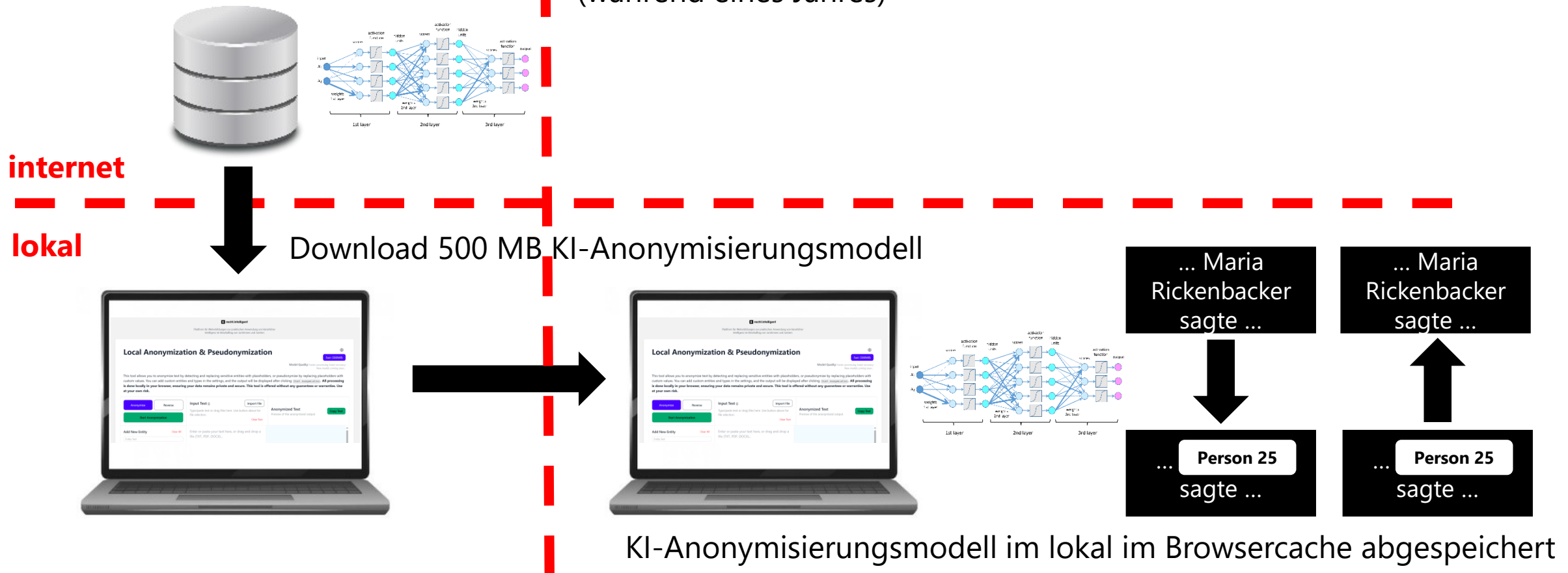
- 1) (Text digitalisieren)
 - 2) Text markieren und kopieren
 - 3) Text anonymisieren
 - Kontrolle !
 - 4) Kommerzielle LLM-Verarbeitung mit anonymisiertem Text
 - 5) LLM-Verarbeitung de-anonymisieren
 - 6) LLM-Verarbeitung verwenden
- lokale Anonymisierung**
- externe Verarbeitung mit anonymisierten Daten**
- lokale De-Anonymisierung**



Ansatz 2: <https://iusable.ai>

1. Aufruf im Web

ab 2. Aufruf im Browser
(während eines Jahres)



Ansatz 2: Daten anonymisieren

Praxisbeispiel mit

1. iusable.ai (lokale Anonymisierung) +
2. persönlicher Prompt- und Vorlagenbibliothek +
3. kommerziellem LLM-Provider (bspw. Google Gemini)

Fazit Ansatz 1

+

- in Textverarbeitung integrierte Anwendung möglich
- Sehr datenschutz- und amtsgeheimnis"robust"

-

- Etablierung herausforderungsreich
- lokale LLMs stehen bezüglich "kognitiver Leistungsfähigkeit" den kommerziellen Modellen nach wie vor deutlich zurück (Ladezeiten, Qualität) - benötigen aufwendiges "prompt engineering"

Fazit Ansatz 2



- Nutzt volle "kognitive Leistungsfähigkeit" kommerzieller KI-Modelle
- Etablierung und Umsetzung einfacher als bei lokalen KI-Modellen



- Gefahren und Risiken bezüglich Datenschutz und Geheimniswahrung
- Wenig intuitiver "workflow" mit Systembrüchen (Anwendungswechsel)
- Wartezeiten und Kontrollaufwand bei Anonymisierung

(Ansatz 3)

- Einverständnis von Kunden/Klienten zur Verarbeitung ihrer persönlicher Daten einholen und/oder
- Auf Versprechen/Zusicherung vertrauen, dass der LLM-Provider die übermittelten Daten nicht verwenden/verwerten wird:
 - [lumo by Proton](#)
 - [ALPINEAI SwissGPT Gov](#)
 - [LexCube](#)
 - [Opt-Out-Vereinbarung](#) bei kommerziellem Provider

"prompt- und context-engineering"

"*prompt*" = An das LLM gerichtete schriftliche Aufgabenstellung

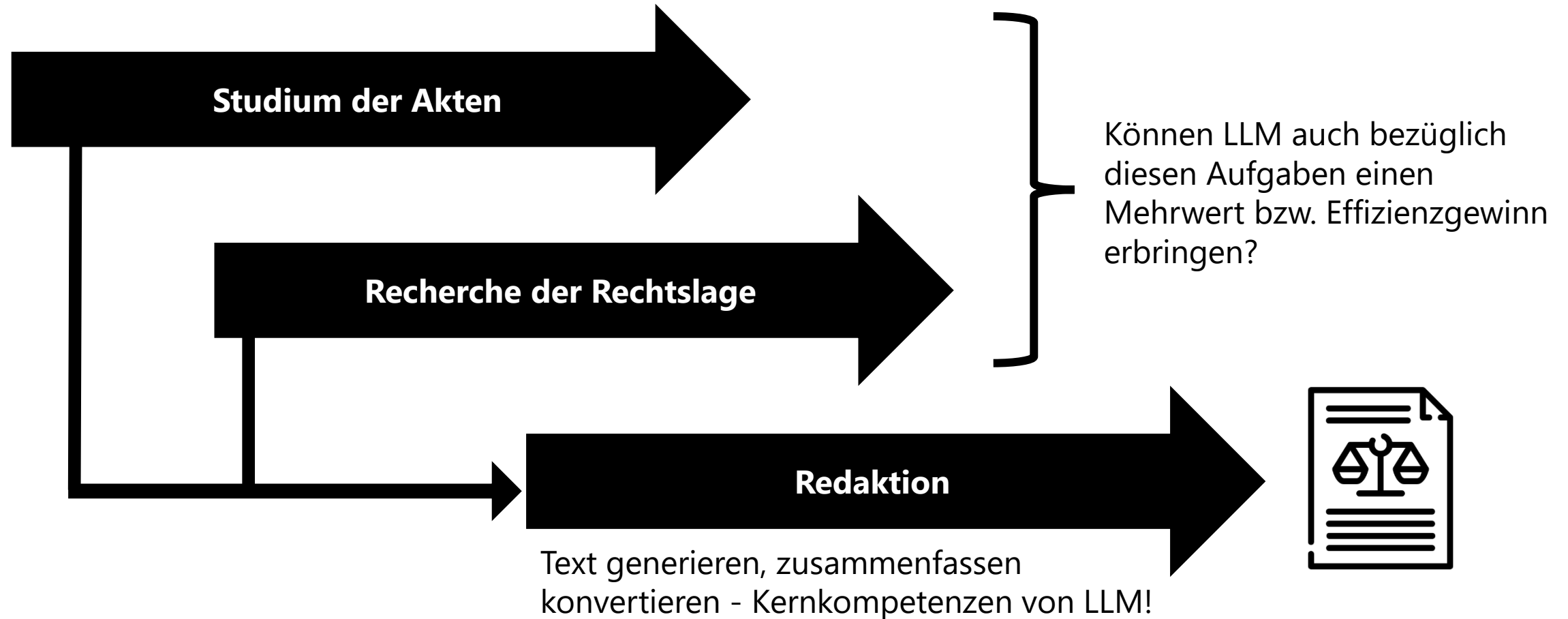
"*context*" = Beispiel(e), die der Aufgabenstellung mitgegeben werden

"*engineering*" = Das Verfassen einer Aufgabenstellung, welche die gewünschten Ergebnisse erbringt

- dem LLM im prompt präzise Anweisungen und Beispiele mitgeben, verbessert die Ergebnisse erheblich
- für das effiziente Arbeiten ist eine "prompt-Bibliothek" unerlässlich
- ausgeklügelte prompts können unter Zuhilfenahme von LLM erstellt werden
- kommerzielle LLM-Provider (mit grossem "context window") erlauben es, dem prompt mehrere grosse Beispieltexte (bspw. anonymisierte Klageschriften) mitzugeben



typische Arbeitsschritte von JuristInnen



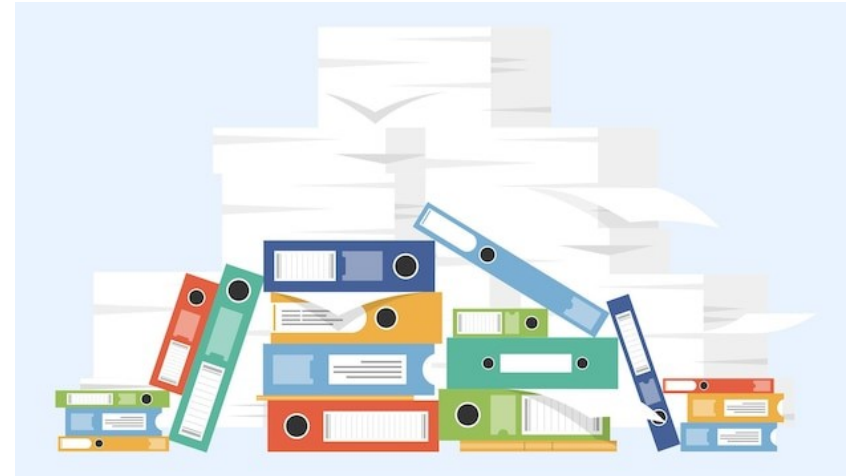
Aktenrecherche

Wie bekomme ich die Informationen aus den Akten in das LLM?

Wie kann ich Ihnen heute helfen?

+

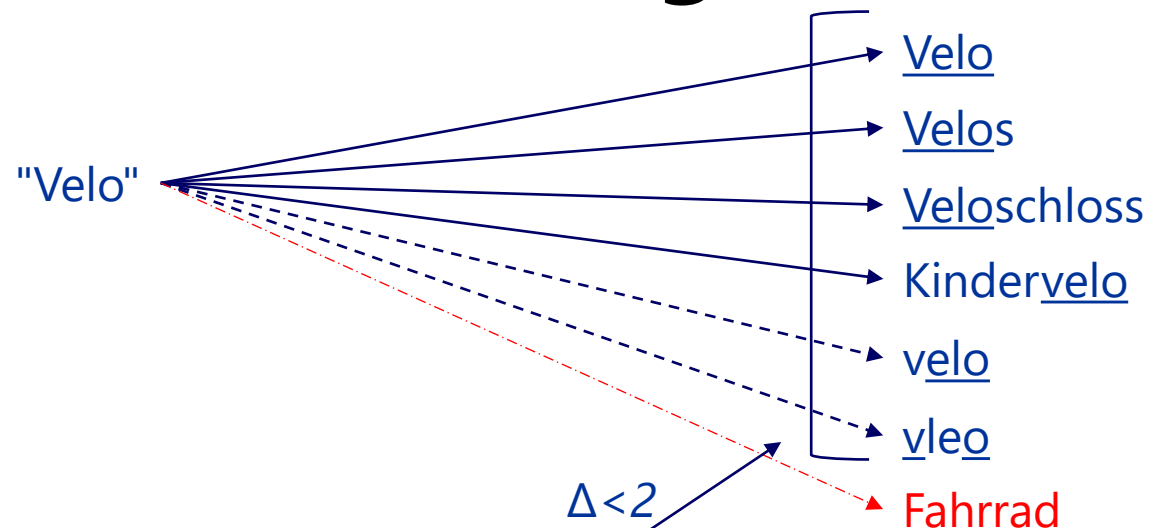
 



Aktenrecherche: "semantic search" bzw. Text-Embedding

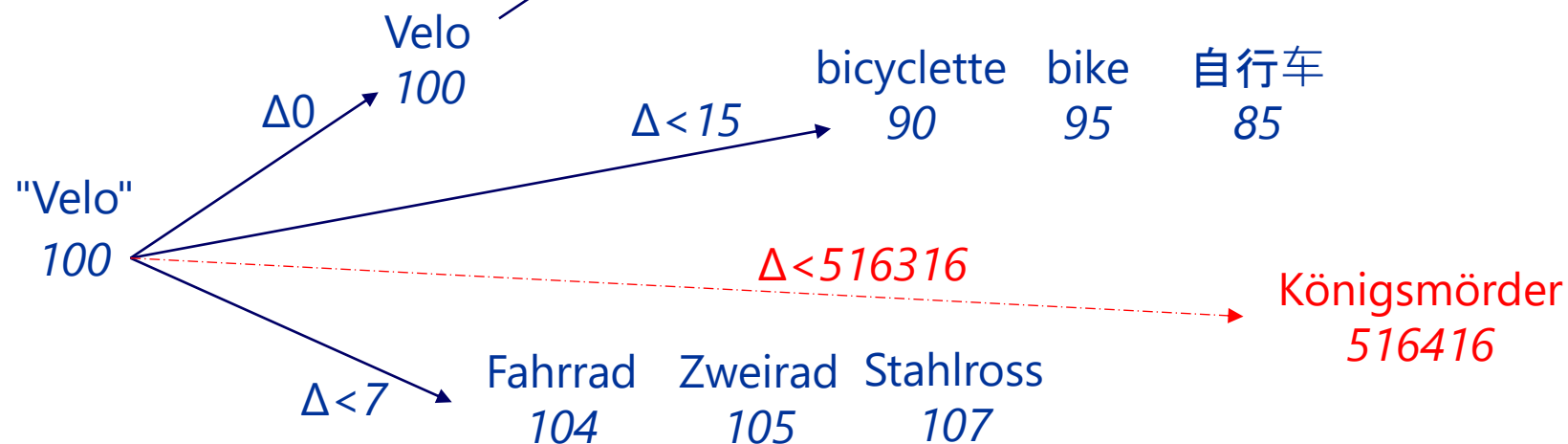
**konventionelle
Suchalgorithmen
"lexical search"**

zeichenbasiert

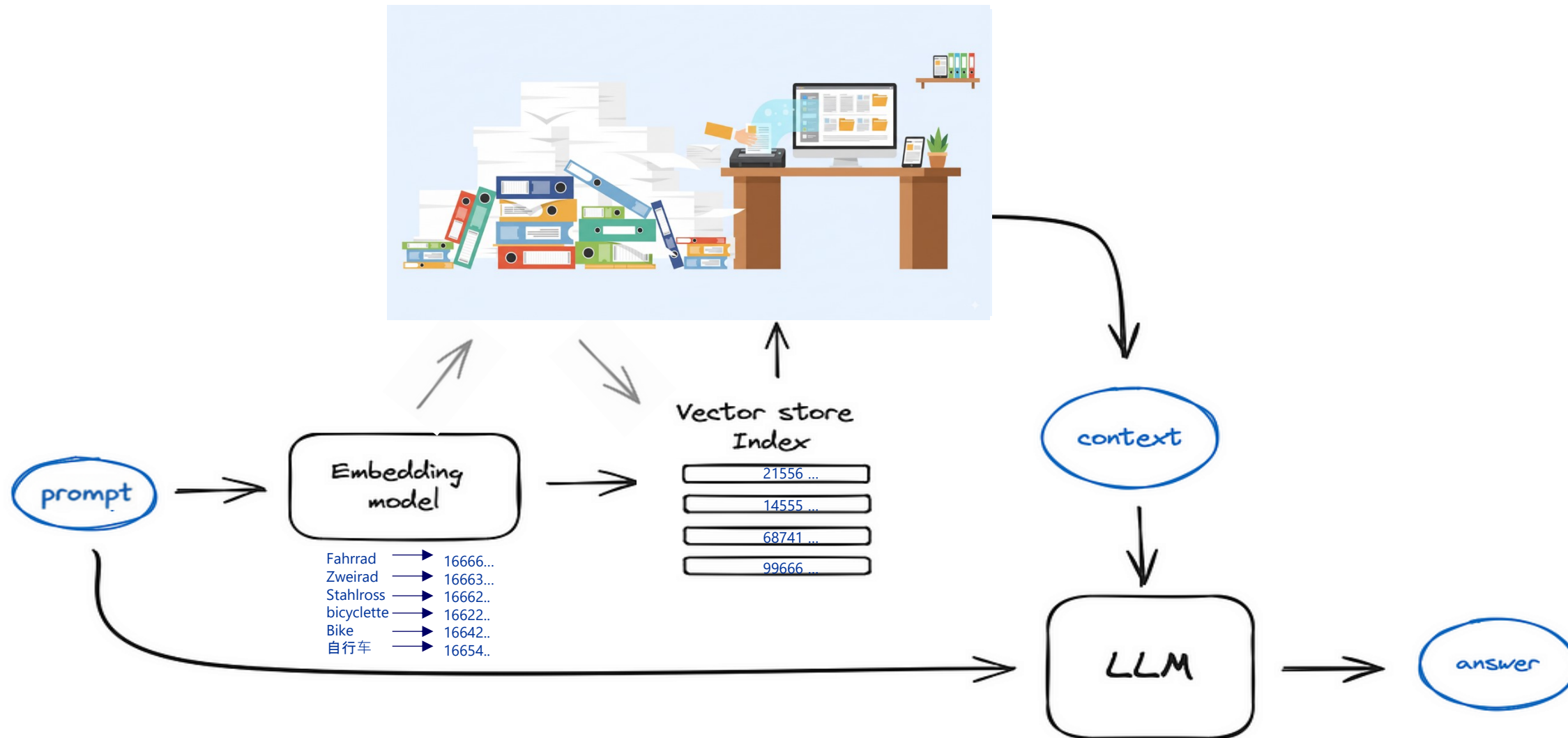


**"semantic search"
Suchalgorithmen**

zahlenbasiert



Aktenrecherche: "Retrieval-augmented Generation" (RAG)



Aktenrecherche: RAG Anwendungen in der Praxis



Anwendungsbeispiel
[FRAG MEISTER JURISTEN](#)

opensource Anwendungen

- [Loom](#) (CH-Armee)
- [machinelearningZH/document-research-tool](#) (Kt. ZH)
- [LangChain](#)
- [Llamaindex](#)
- ...

CH-"Start Ups"

- [DeepJudge](#)
- [Herlock.ai](#)
- [Lexity](#)
- [Casus](#)
- ...

Kommerzielle Anbieter

- [Google NotebookLM](#)
- [OpenAI ChatGPT Projects](#)
- [Google Gemini Gems](#)
- ...

OCR-Problem !

kompliziert

teuer

datenschutzkonformer Betrieb
kaum realisierbar

Aktenrecherche: Praxisbeispiel

Praxisbeispiel

behindertengerechtes Bauen in der Stadt Luzern
(Google NotebookLM)

Rechtsrecherche: Achtung Halluzination!

AI Hallucination Cases

This database tracks legal *decisions*¹ in cases where generative AI produced hallucinated content – typically fake citations, but also other types of arguments. It does not track the (necessarily wider) universe of all fake citations or use of AI in court filings.

While seeking to be exhaustive (2 cases identified so far), it is a work in progress and will expand as new examples emerge. This database has been featured in news media and online posts.² (Readers may also be interested in [this project](#) regarding AI use in academic papers.)

If you know of a case that should be included, feel free to [contact me](#).

<https://www.damiencharlotin.com/hallucinations>

NEWS AI POLICY

Errors found in US judge's withdrawn decision stink of AI



Image: Cath Virginia / The Verge, Getty Images

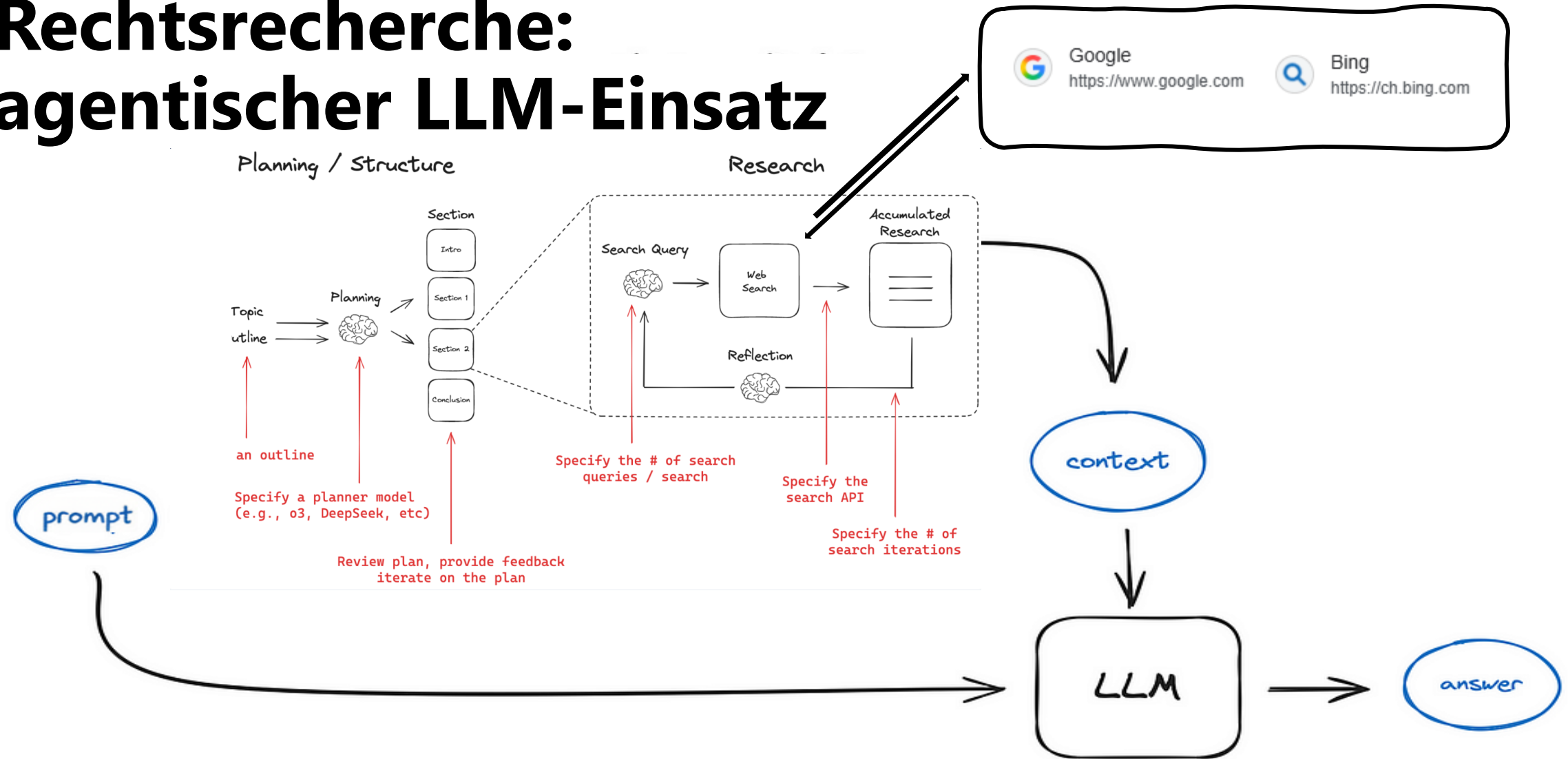
/ The opinion cited made-up quotes and misstated case outcomes, errors that AI tools are known to hallucinate.

by [Jess Weatherbed](#)
Jul 25, 2025, 12:30 PM GMT+2

[Link](#) [Facebook](#) [Twitter](#) [Comments](#)

<https://www.theverge.com/news/713653/judge-withdraws-cormedix-case-ai-citation-errors>

Rechtsrecherche: agentischer LLM-Einsatz



Rechtsrecherche: Angebote

opensource Anwendungen

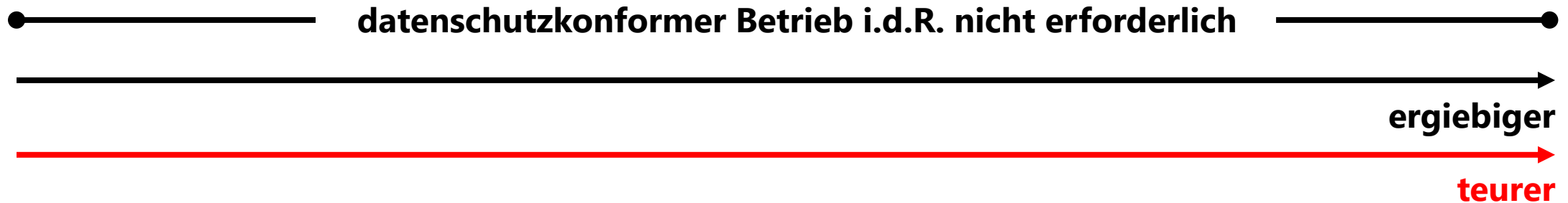
- [machinelearningZH/deep-research](#) (kt. ZH)
- ...

Kommerzielle Anbieter

- [Perplexity](#)
- [OpenAI deep research](#)
- [Google Gemini Deep Research](#)
- ...

CH-"Start Ups"

- [Omnilex](#)
- [Silex](#)
- [Lexi Search](#)
- [Explorer](#)
- ...



Rechtsrecherche: Praxisbeispiel

Praxisbeispiel

behindertengerechtes Bauen in der Stadt Luzern
(Google Gemini Deep Search)

Jonas Achermann

Poolrichter

KANTON LUZERN

Kriminalgericht Luzern

Landenbergstrasse 36

Postfach 3439

6002 Luzern

jonas.achermann@lu.ch

Vertiefungen/Hinweise

recht.intelligent Video-Podcasts:

- [#10 über die Einführung von iusable.ai \(ab Minute 47\)](#)
- [#11 über KI-Opensource-KI-Tools Kt. ZH](#)
- [#12 Red Ink Word KI-Integration Vischer/David Rosenthal](#)

Technische Vertiefungen (englisch):

- [Wie ein neuronales Netzwerk mathematisch funktioniert](#)
- [Wie ein Large Language Modell mathematisch funktioniert](#)
- [Wie ein Large Language Modell funktioniert \(deep dive\)](#)

"prompt engineering":

[OpenAI prompt engineering Erklärvideos \(englisch\)](#)